Econometric Analysis

Ch.2 Conditional Expectation and Related Concepts in Econometrics

Ryuichi Tanaka

2.1 The Role of Conditional Expectations in Econometrics

- Empirical goal: What is the causal relationship between the two events of interest?
- More specific goal: want to know the conditional expectation (CE)
- If we know the CE, we can determine whether a change in one variable, say x, causes a change in another variable, say y
- Definition of CE? Characteristics of CE?

2.2.1 Definition and Examples

- y : dependent variable (random variable drawn from the population)
- $\mathbf{x} \equiv (x_1, \cdots, x_K)$: independent variable (random vector)
- Conditional Expectation

If $E(|y|) < \infty$, then there is a function $\mu : R^K \to R$ such that

$$E(y \mid x_1, \cdots, x_K) = \mu(x_1, \cdots, x_K)$$

$\mu(x_1, \cdots, x_K)$ is a function that shows how a change in x changes y in a causal sense

ex. $E(wage \mid educ, exper, IQ)$

Parametric model (parameterize the CE)

- Candidates of this function are infinite. To make it estimable, we set restriction that the function can be described with finite numbers of parameters
- Example 2.1:

Linear model $E(y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

nonlinear in x, linear in parameters $E(y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2$

nonlinear in x, linear in the betas $E(y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$

nonlinear in the betas $E(y \mid x_1, x_2) = \exp[\beta_0 + \beta_1 \log(x_1) + \beta_2 x_2],\ y \geq 0,\ x_1 > 0$

2.2.2 Partial Effects, Elasticities, and Semielasticities

- Partial Effect Assume that $\mu(x_1, \cdots, x_K)$ is differentiable and x_j is a continuous variable. Then the partial derivative w.r.t. x_j allows us to approximate the marginal change in the CE of y:

$$\Delta E(y|\mathbf{x}) \approx \frac{\partial \mu(\mathbf{x})}{\partial x_j} \Delta x_j \text{ holding } (x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_K) \text{ fixed}$$

Ex. How much does one-year increase in education raise annual labor earnings?

- Interpretation of the magnitudes of coefficients in parametric models comes from the above approximation

Example 2.1

$$E(y|x_1,x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \qquad \frac{\partial E(y|x)}{\partial x_1} = \beta_1 \qquad \frac{\partial E(y|x)}{\partial x_2} = \beta_2$$

$$E(y|x_1,x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 \qquad \frac{\partial E(y|x)}{\partial x_1} = \beta_1 \qquad \frac{\partial E(y|x)}{\partial x_2} = \beta_2 + 2\beta_3 x_2$$

$$E(y|x_1,x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \qquad \frac{\partial E(y|x)}{\partial x_1} = \beta_1 + \beta_3 x_2 \qquad \frac{\partial E(y|x)}{\partial x_2} = \beta_2 + \beta_3 x_1$$

$$E(y|x_1,x_2) = \exp[\beta_0 + \beta_1 \log(x_1) + \beta_2 x_2]$$

$$\frac{\partial E(y|x)}{\partial x_1} = \exp(.)\beta_1 / x_1 \qquad \frac{\partial E(y|x)}{\partial x_2} = \exp(.)\beta_2$$

Elasticity

Elasticity: the percentage change in y when x_j increases by 1 percent

$$\frac{\Delta y / y}{\Delta x_j / x_j}$$

Partial elasticity: the percentage change in the CE of y when x_j increases by 1 percent, ceteris paribus:

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} \frac{x_j}{E(y|\mathbf{x})} = \frac{\partial \mu(\mathbf{x})}{\partial x_j} \frac{x_j}{\mu(\mathbf{x})} \text{ holding } (x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_K) \text{ fixed}$$

Example 2.1 (continued)

$$E(y|x_1,x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \qquad \frac{\partial E(y|\mathbf{x})}{\partial x_1} \frac{x_1}{E(y|\mathbf{x})} = \frac{\beta_1 x_1}{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

$$E(y|x_1,x_2) = \exp[\beta_0 + \beta_1 \log(x_1) + \beta_2 x_2] \qquad \frac{\partial E(y|\mathbf{x})}{\partial x_1} \frac{x_1}{E(y|\mathbf{x})} = \beta_1$$

If $E(y|\mathbf{x}) > 0$ and $x_j > 0$, the equation for the elasticity is the same as $\dfrac{\partial \log[E(y|\mathbf{x})]}{\partial \log(x_j)}$

Semi-elasticity

The percentage change in the CE of y when x_j is increased by one unit is approximated as

$$100 \bullet \frac{\partial E(y\,|\,\mathrm{x})}{\partial x_j} \frac{1}{E(y\,|\,\mathrm{x})} \approx 100 \bullet \frac{\partial \log\left[E(y\,|\,\mathrm{x})\right]}{\partial x_j} \quad \text{if } E(y\,|\,\mathrm{x}) > 0$$

2.2.3 The Error Form of Models of CE

Error Form: y can be decomposed into the sum of the CE of y conditional on $\mathbf{x}$ and the error term with conditional mean zero u, $E(u\,|\,\mathbf{x}) = 0$. That is:

$$y = E(y\,|\,\mathbf{x}) + u,\ E(u\,|\,\mathbf{x}) = 0$$

Two implications of $E(u\,|\,\mathrm{x}) = 0$

(1) $E(u) = 0$ (by LIE)

(2) u is uncorrleated with any function of $\mathbf{x} = (x_1, \cdots, x_K)$,

in particular, u is uncorrelated with each of $(x_1, \cdots, x_K)$

2.2.4 Some properties of CE: Law of Iterated Expectation (LIE)

$\mathbf{w}$: a random vector

y : a random variable

$\mathbf{x}$: a random vector that is some function of $\mathbf{w}$, say, $\mathbf{x} = \mathbf{f}(\mathbf{w})$

If we know the outcome of w, then we know the outcome of x (x is a smaller information set than w)
(x can be just a subset of of w)

(LIE1) $E(y\,|\,\mathbf{x}) = E(E(y\,|\,\mathbf{w})\,|\,\mathbf{x})$

(LIE2) $E(y\,|\,\mathbf{x}) = E(E(y\,|\,\mathbf{x})\,|\,\mathbf{w})$

LIE

(LIE1) $E(y\,|\,\mathbf{x}) = E(E(y\,|\,\mathbf{w})\,|\,\mathbf{x})$

y: earnings w: the name of the last school x: years of education

Want to know the effect of the last school attended on earnings $E(y\,|\,\mathbf{w})$ but if only years of education is available, we can know only the effect of years of education on earnings $E(y\,|\,\mathbf{x})$

(LIE2) $E(y\,|\,\mathbf{x}) = E(E(y\,|\,\mathbf{x})\,|\,\mathbf{w})$

The name of the last school w is redundant for the purpose to estimate the effect of the years of education on earnings $E(y\,|\,\mathbf{x})$

How to memorize the LIEs

"The smaller information set always dominates."

x is a smaller information set than w (if we know the outcome of w, then we know the outcome of x)

If you know the name of the last school w, you know the years of education x. However, even if you know the years of education, you never know the name of the last school without further information

Special case of LIE: $E(y|\mathbf{x}) = E(E(y|\mathbf{x},\mathbf{z})|\mathbf{x})$

x is observable but **z** is not
The above equation relates the CE of **y** when **z** is observable to the CE of y when z is unobservable

$$E(y|x_1,x_2,z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 z$$

When z is unobservable, $E(y|x_1,x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 E(z|x_1,x_2)$

Assume $E(z|x_1,x_2) = \delta_0 + \delta_1 x_1 + \delta_2 x_2$

$$E(y|x_1,x_2) = (\beta_0 + \beta_3\delta_0) + (\beta_1 + \beta_3\delta_1)x_1 + (\beta_2 + \beta_3\delta_2)x_2$$

$\beta_3\delta_i$ omitted variable bias when z is unobservable

Condition by functions

Suppose that for some (vector) function $\mathbf{f}(\mathbf{x})$ and a real-valued function $g(.)$, $E(y|\mathbf{x}) = g(\mathbf{f}(\mathbf{x}))$. Then

$$E(y|\mathbf{f}(\mathbf{x})) = E(y|\mathbf{x}) = g(\mathbf{f}(\mathbf{x}))$$

Example 2.3 $\mathbf{x} = (educ, exper)'$ $\mathbf{f}(\mathbf{x}) = (educ, exper, exper^2, educ*exper)$

$$g(\mathbf{f}(\mathbf{x})) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 educ*exper$$

If $E(wage|educ,exper) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 educ*exper$,

then $E(wage|educ,exper,exper^2,educ*exper) = E(wage|educ,exper)$

Once you take the expectation conditional on *educ* and *exper, you do not have to take expectation conditional on exper^2* or *educ*exper. This implies that taking expectation conditional on the original variables is enough even if you include the functions of these variables*

2.2.5 Average Partial Effects

When the Partial Effect depends on unobserved heterogeneity(q),

$$\frac{\partial E(y|\mathbf{x},q)}{\partial x_j} \equiv \theta_j(\mathbf{x},q)$$

Average Partial Effect (evaluated at $\mathbf{x}^0$): $E_q[\theta_j(\mathbf{x}^0,q)] \equiv \delta_j(\mathbf{x}^0)$

In general, q (and its distribution under x^0) is unknown,

and thus $\delta_j(x^0)$ is not estimable.

A special case where APE is estimable

Assumption 1 : $\mathbf{x}$ and q are independent conditional on $\mathbf{w}$

($\mathbf{w}$ is a proxy variable for q)

$D(q \mid \mathbf{x}, \mathbf{w}) = D(q \mid \mathbf{w})$ ($\mathbf{w}$ is enough to know the distribution of q)

$D(\cdot \mid \cdot)$: conditional distribution

Assumption 2 : $\mathbf{w}$ is *redundant* (*ignorable*) in the structural expectation

$E(y \mid \mathbf{x}, q, \mathbf{w}) = E(y \mid \mathbf{x}, q)$ ($\mathbf{x}$ and q are enough to know the CE of y)

Under the two assumptions, $\delta_j(\mathbf{x}^0) = E_{\mathbf{w}}[\partial E(y \mid \mathbf{x}^0, \mathbf{w}) / \partial x_j]$

2.3 Linear Projection

$(y, x_1, \cdots, x_K)$:random variables with $E(y^2) < \infty$, $E(x_j^2) < \infty$

$\mathbf{x} \equiv (x_1, \cdots, x_K)$: 1 x K vector, where the K x K variance matrix of x is nonsingular (positive definite)

Linear projection of y on (1,x)

Under the above assumptions, there always exists a unique linear projection of y on (1,x) such that

$$L(y \mid 1, x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K = \beta_0 + \mathbf{x}\boldsymbol{\beta}$$

$$\beta = [Var(x)]^{-1} Cov(x, y) \qquad \beta_0 = E(y) - E(x)\beta$$

LP

Note parameters in LP are defined in terms of the population moment (mean, variance, etc.)

OLS replaces these moments by sample analog to estimate the parameters

Error Form of LP

LP can be written as an error form

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u, \quad E(u) = 0 \quad Cov(x_j, u) = 0$$

Problem Set 1

1. 2.1
2. 2.2
3. 2.3
4. 2.4
5. 2.6